

## INTISARI

### PERBANDINGAN PENGGUNAAN *WORD EMBEDDING* WORD2VEC, GLOVE, DAN FASTTEXT DALAM PERINGKASAN TEKS EKSTRAKTIF BERITA BERBAHASA INDONESIA MENGGUNAKAN RNN

Oleh

Zahra Khaerunnisa

18/433805/PPA/05620

Beberapa tahun terakhir pendekatan dengan metode *Deep Neural Network* telah banyak digunakan pada peringkasan teks ekstraktif untuk dapat meningkatkan performa secara signifikan. Namun tidak ada pemahaman yang jelas terkait bagaimana metode tersebut dapat bekerja dengan baik ketika digunakan pada sistem. Salah satu alasan peningkatan akurasi dan kinerja pada model *Deep Neural Network* terjadi adalah penggunaan pengetahuan eksternal yaitu *pre-trained word embedding*, representasi kata dalam bentuk vektor yang telah dilatih menggunakan dataset yang besar pada domain permasalahan tertentu. Jenis *word embedding* yang sering digunakan adalah *unsupervised learning* diantaranya Word2Vec, GloVe, dan FastText yang mempunyai mekanisme berbeda untuk menghasilkan representasi kata dalam bentuk vektor. Tetapi diantara *word embedding* tersebut belum diketahui pasti jenis mana yang lebih baik digunakan untuk peringkasan teks ekstraktif. Metode *Deep Neural Network* telah banyak digunakan untuk menyelesaikan permasalahan peringkasan teks ekstraktif salah satunya penggunaan *Recurrent Neural Network* (RNN) dengan *Gated Recurrent Unit* (GRU) yang digunakan sebagai ekstraksi fitur dan sebagai klasifikasi untuk kalimat ringkasan.

Penelitian ini dilakukan untuk membandingkan kinerja dari pada peringkasan teks ekstraktif yang menggunakan jenis *word embedding* Word2Vec, GloVe, dan FastText dan *Recurrent Neural Network* dengan *Gated Recurrent Unit* digunakan sebagai klasifikasi kalimat ringkasan pada berita bahasa Indonesia. Evaluasi model diukur menggunakan F-measure metrik ROUGE-1, ROUGE-2, dan ROUGE-L.

Performa terbaik dihasilkan oleh model yang menggunakan *word embedding* FastText dengan nilai F-Measure ROUGE-1 70.82, ROUGE-2 66.525 dan ROUGE-L 70.288.

**Kata Kunci:** Peringkasan Teks Ekstraktif, *Word Embedding*, Word2Vec, GloVe, FastText, *Recurrent Neural Network*

## ABSTRACT

### THE COMPARISON OF WORD2VEC, GLOVE, AND FASTTEXT WORD EMBEDDING IN EXTRACTIVE TEXT SUMMARIZATION OF INDONESIAN NEWS BY USING RNN

by

Zahra Khaerunnisa

18/433805/PPA/05620

In recent years, the Deep Neural Network approach has been widely used in extractive text summarization to significantly improve performance. However, there is no clear understanding of how this method can work properly when used on the system. One of the reasons for increasing accuracy and performance in the Deep Neural Network model is the use of external knowledge, namely pre-trained word embedding, word representation in vector form that has been trained using large datasets in certain problem domains. The type of word embedding that is often used is unsupervised learning including Word2Vec, GloVe, and FastText which have different mechanisms to produce word representations in vector form. However, among the word embedding, it is not clear which type is better to use for extractive text summarization. The Deep Neural Network method has been widely used to solve extractive text summarization problems, one of which is the use of Recurrent Neural Network (RNN) with Gated Recurrent Unit (GRU) which is used as feature extraction and as a classification for summary sentences.

This research was conducted to compare the performance of extractive text summarization using Word2Vec, GloVe, and FastText word embedding and Recurrent Neural Network with Gated Recurrent Unit used as summary sentence classification in Indonesian news. Evaluation of the model was measured using the F-measure metrics ROUGE-1, ROUGE-2, and ROUGE-L.

The best performance was generated by the model using word embedding FastText with F-Measure values ROUGE-1 70.82, ROUGE-2 66.525 and ROUGE-L 70.288.

**Keywords:** *Extractive Text Summarization, Word Embedding, Word2Vec, GloVe, FastText, Recurrent Neural Network*